10/521877
DT12 Rec'd PCT/PTO 20 JAN 2005

Claims

Claim 1, the characteristics of the RDO based rate control scheme include:

Step 1: Does bit allocation for every picture in a GOP, and based on the allocated bits a predicted quantization parameter is used to do rate distortion optimization mode selection for every macroblock in the current picture;

Step 2: The information collected from the first rate distortion mode selection is used to calculate a final quantization parameter for rate control, and if the final quantization parameter is different from the predicted, a second rate distortion mode selection will be executed again.

Claim 2, as claim 1 has said, in step 1, before coding a GOP, does bit allocation for the pictures in the GOP with the average picture size;

Claim 3, as claim 2 has said, the average picture size is calculated as:

$R/F = R \div F$, here, R is the target bit rate. F is the picture rate. R/F is the average picture size.

Claim 4, as claim 1 and claim 2 have said, does bit allocation adjustment in the coded GOP. The adjustment is implemented as follows:

$$T_i = \max\left\{ \frac{R}{1 + \frac{N_p X_p}{K_p X_i} + \frac{N_b X_b}{K_b X_i}}, \frac{bit\_rate}{8 \times picture\_rate} \right\}$$

$$T_p = \max\left\{ \frac{R}{N_p + \frac{N_b K_p X_b}{K_b X_p}}, \frac{bit\_rate}{8 \times picture\_rate} \right\}$$

$$T_b = \max\left\{ \frac{R}{N_b + \frac{N_p K_b X_p}{K_p X_b}}, \frac{bit\_rate}{8 \times picture\_rate} \right\}$$

here, T_I, T_p and T_b is the bits allocated to the I, P or B frame respectively. N_i, N_p and N_b is the remained none coded I, P or B frames in the GOP respectively. X_i, X_p and X_b is the global complexity estimation for the I, P or B frame respectively and is defined as the multiplier between coded bits and average quantization parameter for the frame.

bit_rate is the target bit rate. *picture_rate* is the frame rate.

K_p and K_b are constants. K_p, K_b means the complexity ratio between P, B frame and I frame respectively.

R is the remained bits for the GOP, and after coding a picture it is updated as follows:

$R = R\text{-}S_{i,p,b}$

$S_{i,p,b}$ is the coded bits for the current frame.

Claim 5, as claim 4 has said, before coding a GOP, the remaining bits for the current GOP is initialized as follows:

$$R = G + R_{prev}$$

$$G = bit\_rate \times N \div picture\_rate$$

here, R is the remained bits for the current GOP.

N is the number of frames in the current GOP.

G is the number of bits for a GOP.

R_{prev} is the remained bits for the previous GOP. For the first GOP, R_{prev}=0.

Claim 6, as claim 4 has said, X_i、 X_p and X_b are initialized as:

$$X_i = a \times bit\_rate$$

$$X_p = b \times bit\_rate$$

$$X_b = c \times bit\_rate$$

here a, b and c are constants.

bit_rate is the target bitrate.

Claim 7, as claim 1 has said, the step 1 also includes at least one time rate distortion optimization based mode selection with a predicted quantization parameter. The predicted quantization parameter may be the quantization parameter of the previous macroblock or decided by rate distortion model in a rate control scheme. The mode minimizing the following expression is selected as the initial coding mode for the current macroblock:

$$D(s,c,MODE \mid QP) + \lambda_{MODE} R(s,c,MODE \mid QP)$$

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_{m-1}/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_{m-1}/3}$.

D(s,c,MODE|QP) is used to evaluate the distortion of the current macroblock after it is coded with mode *MODE.*

R(s,c,MODE|QP) is the bits used to code the macroblock with mode *MODE.*

QP is the quantization parameter for the current macroblock.

Claim 8, as claim 7 has said, for motion estimation in P or B frame, the motion vector minimizing following expression is selected as the motion vector for the current macroblock:

$$J(\mathrm{m}, \lambda_{MOTION}) = SA(T)D(s, c(\mathrm{m})) + \lambda_{MOTION} R(\mathrm{m} - \mathrm{p})$$

here, *D(s,c(m))* is used to evaluate the distortion from motion compensation.

SA (T) D is the sum of the absolute difference after prediction (or after Hadmard transform) for the macroblock.

R(m-p) is the bits used to code the motion vector.

s is the luma value of the current macroblock in the original frame.

c is the luma value in reference picture.

m is the motion vector.

p is the predicted motion vector.

λ_{MOTION} is the lagrangian constant and $\lambda_{MOTION} = \sqrt{\lambda_{MODE}}$.

λ_{MODE} is the lagrangian constant.

Claim 9, as claim 2 has said, after the first rate distortion mode selection, the RDO based rate control further includes: calculating quantization parameter for the current macroblock. The quantization parameter is adjusted according to the macroblock activity and buffer status.

Claim 10, as claim 9 has said, the quantization parameter for the macroblock is adjusted according to the macroblock activity. After the first rate distortion mode selection, the sum of the absolute difference is used as the macroblock activity estimation. The macroblock activity is calculated as:

$$act_m = \sum_{i,j} |s(i,j) - c(i,j)| \quad N\_act_m = \frac{(2 \times act_j) + avg\_act}{act_j + (2 \times avg\_act)}$$

here, i is the horizontal position of the pixel in the current macroblock. j is the vertical position of the pixel in the current macroblock. N_act_m is the activity of the current macroblock. $s(i,j)$ is the luma value of the original pixel(i,j), $c(i,j)$ is the prediction value of pixel(i,j). avg_act is the average act_m in the previous coded picture which is coded with the same type as current picture. act_m is the sum of the absolute difference after motion compensation or intra prediction.

Claim 11, as claim 9 has said, a virtual buffer is used to do rate control. First set up the mapping from the virtual buffer occupancy to macroblock quantization parameter, and the final macroblock quantization parameter is calculated as:

$$Q_m = \left(\frac{d_m^n \times 31}{r}\right) \times N\_act_m$$

$$d_m^n = d_0^n + B_{m-1} - T_n \times (m-1) / MB\_CNT$$

$$r = 2 \times bit\_rate / picture\_rate$$

here, Q_m is the quantization parameter of current macroblock.

d_m^n is the current buffer occupancy, and it equals d_m^I, d_m^P, and d_m^b for I, P, B frame respectively.

B_{m-1} is the bits used to code previous macroblock.

d_0^n is the initial buffer occupancy for current frame. n is i,p or b, corresponding to d_0^i, d_0^P,and d_0^b.

r is the size of virtual buffer.

Claim 12, as Claim 11 said, when coding the first frame, the virtual buffer occupancy is initialized with:

$$d_0^b = K_b \times d_0^i$$

$$d_0^i = 10 \times r / 31$$

$$d_0^p = K_p \times d_0^i$$

here r is the virtual buffer size; d_0^i, d_0^P, and d_0^b is the initial virtual buffer occupancy for i, p, or b frame. K_p is the complexity ratio between I, P frame; K_b is the complexity ratio between I, B frame.

Claim 13, as claim 2, 9, 10, 11 and 12 have said, the RDO based rate control also includes a

second RDO mode selection, after calculating the final quantization parameter for the current macroblock. That is to say, the selected quantization parameter for the current macroblock will be used to do RDO mode selection again. The mode which minimizes the following expression will be selected as the coding mode for the current macroblock:

$$D(s,c,MODE \mid QP) + \lambda_{MODE} R(s,c,MODE \mid QP)$$

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_{m-1}/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_{m-1}/3}$.

D(s,c,MODE|QP) is used to evaluate the distortion of the current macroblock coded with mode *MODE*.

R(s,c,MODE|QP) is the bits used to code the macroblock with mode *MODE*.

QP is the quantization parameter for the current macroblock.

Claim 14, as claim 13 has said, for motion estimation in P or B frame, the motion vector minimizing following expression is selected for the current macroblock:

$$J(m, \lambda_{MOTION}) = SA(T)D(s, c(m)) + \lambda_{MOTION} R(m - p)$$

here, *D(s,c(m))* is used to evaluate the distortion from motion compensation.

SA（T）D is the sum of the absolute difference (or after Hadmard transform) for the macroblock.

R(m-p) is the bits used to code the motion vector.

s is the luma value of the current macroblock in the original frame.

c is the luma value in reference picture.

m is the motion vector.

p is the predicted motion vector.

λ_{MOTION} is the lagrangian constant and $\lambda_{MOTION} = \sqrt{\lambda_{MODE}}$.

λ_{MODE} is the lagrangian constant.

Claim 15, a rate distortion optimization based rate control implementation includes following modules: a video coding encoder module (for example, H.264 encoder module or JVT processing module), rate distortion optimization mode selection and adaptive quantization module,

virtual buffer, and global complexity estimation module; here, JVT processing module receives the input frame, and it is connected with RDO mode selection module, virtual buffer module and global complexity estimation module.

RDO mode selection module and adaptive quantization is connected with virtual buffer and global complexity estimation module. It receives the input signal from JVT processing module, and processes it based on the virtual buffer module and global complexity module status, and then calculate the quantization parameter for the macroblock. In the last, JVT processing module will output the final coded macroblock with the calculated parameter.

Claim 16, as claim 15 has said, before coding a GOP, does bit allocation for the pictures in the GOP with the average picture size;

Claim 17, as claim 16 said, the average picture size is calculated as:

$R/F = R \div F$, here, R is the target bit rate. F is the picture rate. R/F is the average picture size.

Claim 18, as claim 16 and 17 have said, does bit allocation adjustment in the GOP. The adjustment is shown as follows:

$$T_i = \max\left\{\frac{R}{1+\dfrac{N_pX_p}{K_pX_i}+\dfrac{N_bX_b}{K_bX_i}}, \frac{bit\_rate}{8\times picture\_rate}\right\}$$

$$T_b = \max\left\{\frac{R}{N_b+\dfrac{N_pK_bXp}{K_pX_b}}, \frac{bit\_rate}{8\times picture\_rate}\right\}$$

$$T_p = \max\left\{\frac{R}{N_p+\dfrac{N_bK_pX_b}{K_bX_p}}, \frac{bit\_rate}{8\times picture\_rate}\right\}$$

here, T_I, T_p and T_b is the bits allocated tho the I, P or B frame respectively. N_i、N_p and N_b is the remained none coded I, P or B frames in the GOP respectively. X_i, X_p and X_b is the global complexity estimation for the I, P or B frame respectively and is defined as the multiplier between the coded bits and average quantization parameter for the frame.

bit_rate is the target bit rate. *picture_rate* is the frame rate.

K_p and K_b are constants. K_p, K_b means the complexity ratio between P,B frame and I frame

respectively.

R is the remained bits for the GOP, and after coding a picture it is updated as follows:

$R = R\text{-}S_{i,p,b}$

$S_{i,p,b}$ is the coded bits for the current frame.

Claim 19, as Claim 18 has said, before coding a GOP, the remaining bits for the current GOP is initialized as follows:

$$R = G + R_{prev}$$

$$G = bit\_rate \times N \div picture\_rate$$

here, R is the remained bits for the current GOP.

N is the number of frames in current GOP.

G is the number of bits for a GOP.

R_{prev} is the remained bits for the previous GOP. For the first GOP, R_{prev}=0.

Claim 20, as claim 18 said, X_i、 X_p and X_b are initialized as:

$$X_i = \mathrm{a} \times bit\_rate$$

$$X_p = \mathrm{b} \times bit\_rate$$

$$X_b = \mathrm{c} \times bit\_rate$$

here a, b and c are constants.

bit_rate is the target bitrate.

Claim 21, as claim 15 said, does the mode selection while using the quantization parameter of previous macroblock as a prediction value for the current macroblock. The mode minimizing the following expression is selected as the initial coding mode for the current macroblock:

$$D(s,c,MODE \mid QP) + \lambda_{MODE} R(s,c,MODE \mid QP)$$

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_{m-1}/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_{m-1}/3}$.

$D(s,c,MODE|QP)$ is used to evaluate the distortion of the current macroblock coded with mode $MODE$.

$R(s,c,MODE|QP)$ is the bits used to code the macroblock with mode $MODE$.

QP is the quantization parameter for the current macroblock.

Claim 22, as claim 21 has said, for motion estimation in P or B frame, the motion vector minimizing following expression is selected for the current macroblock:

$$J(\mathrm{m}, \lambda_{MOTION}) = SA(T)D(s, c(\mathrm{m})) + \lambda_{MOTION} R(\mathrm{m} - \mathrm{p})$$

here, *D(s,c(m))* is used to evaluate the distortion from motion compensation.

SA (T) D is sum of the absolute difference (or after Hadmard transform) for the macroblock.

R(m-p) is the bits used to code the motion vector.

s is the luma value of the current macroblock in the original frame.

c is the luma value in reference picture.

m is the motion vector.

p is the predicted motion vector.

λ_{MOTION} is the lagrangian constant and $\lambda_{MOTION} = \sqrt{\lambda_{MODE}}$.

λ_{MODE} is the lagrangian constant.

Claim 23, as claim 22 has said, after the first rate distortion mode selection, the rate control scheme further includes: Calculating a new quantization parameter and adjusting it according to the macroblock activity and buffer status.

Claim 24, as claim 22 said, for adjusting quantization parameter for the current macroblock. the sum of the absolute difference is used as the macroblock activity estimation after first rate distortion mode selection. The macroblock activity is calculated as:

$$act_m = \sum_{i,j} |s(i,j) - c(i,j)| \quad N\_act_m = \frac{(2 \times act_j) + avg\_act}{act_j + (2 \times avg\_act)}$$

here, *i* is the horizontal position of the pixel in the current macroblock. *j* is the vertical position of the pixel in the current macroblock. N_act_m is the activity of the current macroblock. *s(i,j)* is the luma value of the original pixel(*i,j*), *c*(*i,j*) is the prediction value of pixel(*i,j*). *avg_act* is the average act_m in the previous coded picture which is coded with the same type as current picture. act_m is the sum of the absolute difference after motion compensation or intra prediction.

Claim 25, as claim 22 has said, a virtual buffer is used to do rate control. First set up the mapping from the virtual buffer occupancy to macroblock quantization parameter. The macroblock

quantization parameter is calculated as:

$$Q_m = \left(\frac{d_m^n \times 31}{r}\right) \times N\_act_m$$

$$r = 2 \times bit\_rate / picture\_rate$$

$$d_m^n = d_0^n + B_{m-1} - T_n \times (m-1) / MB\_CNT$$

here, Q_m is the quantization parameter of current macroblock.

d_m^n is the current buffer occupancy, and it equals d_m^I, d_m^P, and d_m^b for I, P, B frame respectively.

B_{m-1} is the bits used to code previous macroblock.

d_0^n is the initial buffer occupancy for current frame. n is i,p or b, corresponding to d_0^i, d_0^P, and d_0^b.

r is the size of virtual buffer occupancy.

Claim 26, as claim 25 said, when coding the first frame, the virtual buffer occupancy is initialized with:

$$d_0^i = 10 \times r / 31$$

$$d_0^p = K_p \times d_0^i$$

$$d_0^b = K_b \times d_0^i$$

here r is the virtual buffer size; d_0^i, d_0^{P}, and d_0^b is the initial virtual buffer occupancy for i, p, or b frame. K_p is the complexity ratio between I, P frame; K_b is the complexity ratio between I,B frame.

Claim 27, as claim 23, 24, 25, 26 have said, the RDO based rate control also includes a second RDO mode selection, after quantization parameter decision for the current macroblock. That is to say, the decided quantization parameter for the current macroblock will be used to do RDO mode selection again. The mode which minimizes the following expression will be selected as coding mode for the current macroblock:

$$D(s, c, MODE \mid QP) + \lambda_{MODE} R(s, c, MODE \mid QP)$$

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_m - 1/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_m - 1/3}$.

D(s,c,MODE|QP) is used to evaluate the distortion of the current macroblock after it is coded.

R(s,c,MODE|QP) is the bits used to code the macroblock with mode *MODE*.

QP is the quantization parameter for current macroblock.

Claim 28, as claim 27 said, for motion estimation in P or B frame, the motion vectors minimizes following expression are selected as the motion vectors for the current macroblock:

$$J(\mathrm{m}, \lambda_{MOTION}) = SA(T)D(s, c(\mathrm{m})) + \lambda_{MOTION} R(\mathrm{m} - \mathrm{p})$$

here, *D(s,c(m))* is used to evaluate the distortion from motion compensation.

SA (T) D is sum of the absolute difference (or after Hadmard transform) for the macroblock.

R(m-p) is the bits used to code the motion vector.

s is the luma value of the current macroblock in the original frame.

c is the luma value in reference picture.

m is the motion vector.

p is the predicted motion vector.

λ_{MOTION} is the lagrangian constant and $\lambda_{MOTION} = \sqrt{\lambda_{MODE}}$.

λ_{MODE} is the lagrangian constant.

Claim 29, as Claim 28 said, quantization parameter from RDO and adaptive quantization module is sent back to JVT processing module, the macroblock is coded by JVT processing module and output.